Final Report

Title: Chance Discovery with Data Crystallization

Discovering Unobservable Events

AFOSR/AOARD Reference Number: AOARD-06-4062

AFOSR/AOARD Program Manager: **Hiroshi Motoda**, Ph.D.

Period of Performance: 01 Sept 2006 - 1 Sept 2007

Submission Date: 19 November 2007

PI: Katsumi Nitta / Tokyo Institute of Technology

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **12 APR 2007**	2. REPORT TYPE	3. DATES COVERED

4. TITLE AND SUBTITLE **Chance Discovery with Data Crystallization - Discovering Unobservable Events**	5a. CONTRACT NUMBER **FA48690610120**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Katsumi Nitta**	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Tokyo Institute of Technology,4259 Nagatsuta, Midori-ku,Yokohama ,JP,176-0024**	8. PERFORMING ORGANIZATION REPORT NUMBER **N/A**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

he method of data crystallizing reveals the hidden structure by inserting dummy items corresponding to unobservable, i.e., hidden events, to the given incomplete and ill-structured data on past events. The existence of those hidden events and their location in the environment were visualized as a result of data crystallization. The method was evaluated by applying to 1) the simulated data using the 9/11 terrorist network and 2) test data provided by Dr. Bob Schrag via. It was further applied to two real-business domains: 1) redesigning surface inspection system (SIS), and 2) extracting the essence of flow of arguments in the negotiation logs for hypothetical two companies with respect to three business proposals.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **51**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

(2) Objectives: Briefly summarize the objectives of the research effort or the statement of work.
It is only the observable part of the real world that can be presented in data. From such scattered data, i.e., incomplete and ill-structured data, the method of *data crystallizing* presented by Ohsawa in a preceding project (AOARD-05-15) revealed the hidden structure by inserting dummy items corresponding to unobservable, i.e., hidden events, to the given data on past events. The existence of those hidden events and their position in the environment were visualized as a result of data crystallizing. This year, this basic method has been extended to be applicable to various real world domains such as intelligence analysis of terrorist networks, product development, and also to noticing essential hidden assertions in disputes. In this project, the researchers led by new PI Nitta has been developing a human-centric process for improving the performance of data crystallization, with inventing (a) a new tool extending KeyGraph, and (2) a process to involve human's interpretation and the iterative manipulation of the visualized graph. The planned experiments in the proposal were to analyze:

- Artificial data obtained from simulating the target of intelligence analysis, i.e., organized crimes.
- Other kinds of data, which are matching emerging social interests, e.g., the real text on the conversations in the court, and the data on humans' movements in town.

(3) Status of effort: A brief statement of progress towards achieving the research objectives. (Limit this section to about 200 words or less.)

The basic algorithm of *data crystallizin*g has been realized in AOARD-054016, to visualize unobservable events and their relations with other events. This year, new PI Katsumi Nitta succeeded the work above, to organize the project on data crystallization to be a more widely applicable method for visualizing hidden events in the real world. Nitta is progressing this year with the assist of Yoshiharu Maeno, who is a PhD candidate under Ohsawa's supervision. This project finally established a method to discover a node which is significantly relevant to others in a complex social network but missing in the data. The problem was difficult, not only because such a node (corresponding to a leader of terrorist group) appears infrequently and non-routinely, but also because the logs of analysts' surveillance on the covert social network is hardly available. We invented a method for integrating the investigator's prior understanding, insight on the target social network, and computational data processing. We evaluated the method by applying to the simulated data using the 9/11 terrorist network and test data provided by Dr. Bob Schrag via Dr. Tae-Woo Park. We also applied to business problems. Inventing a new idea in corporate research and development is studied.

(4) Abstract: Briefly describe research accomplishments, their significance to the field, and their relationship to the original goals.

Accomplishments

a. *Stage 1) Development of basic tool* : For a scattered, i.e., an incomplete and ill-structured dataset, we realized a tool for *data crystallizing* which inserts dummy items, corresponding to unobservable events. The existence of these unobservable events and their relations with other events are visualized by applying KeyGraph iteratively to the data donated with dummy items, gradually increasing the number of edges in the graph, like the crystallization of snow with gradual decrease in the air temperature. For tuning the granularity level of structure to be visualized, this tool is integrated with human's process of chance discovery. This basic method came to be proven applicable for the discovery of hidden leaders of meetings, i.e., managers who do not appear in the meeting room but are sending commands to the members who appear in the meetings.

b. *Stage 2) Refinement of the method by weighing human's role in the process of discovery* : We addressed hidden structure visualization adaptive to human's prior understanding. Visualization can be adjusted based on the degree of the user's prior understanding of the problem domain. The degree is represented by a temperature parameter used in the human-interactive annealing along with stable deterministic crystallization algorithm. When the understanding of the problem is believed to be richer, the temperature shall be set higher. More complex higher-order hidden structures shall be revealed. This will lead to the discovery of unique and unexpected scenario. On the other hand, when the understanding is poorer, the temperature shall be set lower. The user should try to understand the basic lower-order structures from the event graph. Such adaptive nature is convenient to discover unexpected scenarios in the individual user's own perspective. The adaptive nature of the annealing process was demonstrated for examples of social network visualizations from: (1) Test data generated from a scale-free network, resulting in the discovery precision of up to 90%. (2) Real on-line communication where people met for group decision, resulting in precisely discovering real leaders who had been deleted from the data of communication (3) data of persons related to famous politicians.

c. *Stage 3) Extension of the algorithm and the process, and evaluations for simulated datasets* : The extension is an interactive process starting from the analysts' surveillance, where the hypotheses on the latent structure are discussed. The algorithm, in case of terrorist analysis, visualizes the data on the terrorists' communication and show a social network diagram to the analysts. It consists of clustering and ranking procedure. In the clustering procedure, the activeness of communications between the terrorists are computed and visualized, and uses the analysts' prior knowledge such as the normal number of groups or the known group leaders. The ranking procedure computes the likeliness of inter-cluster relationships, which originates in the unobserved person hidden in the empty spots between the clusters, and indicates the position of the person as a red node. The investigators compare the visualized social network diagram and prior knowledge, and update the prior understanding, iterate the above procedures, and finally invent a hypothesis on the latent structure (Maeno, 2007). The details of the method are presented in the attachment I. The two figures in comparison shows an example of result. Other results shown in the attachment shows the high accuracy of detecting hidden leaders, according to our application to the AOARD test data.

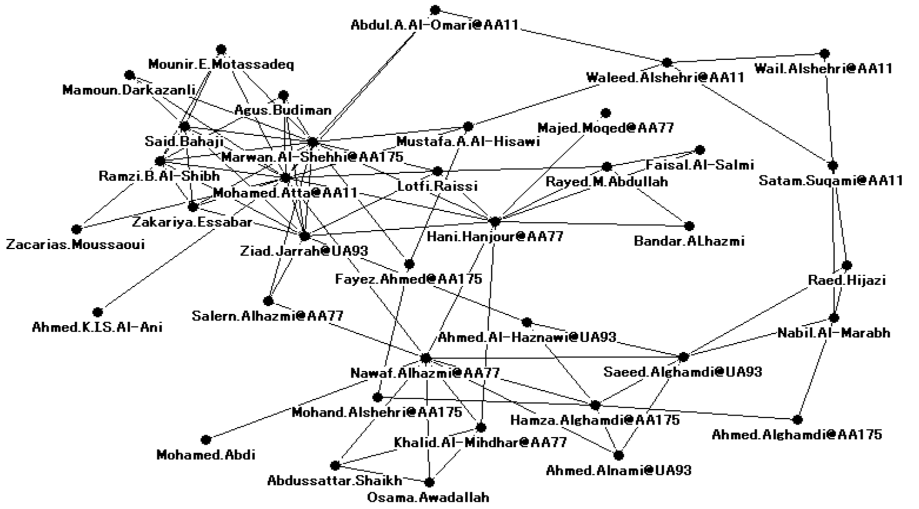

Figure A (corresponding to Fig.4 in attachment I) Social network diagram representing the observed 19 hijackers responsible for the 9/11 attack in figure 3 with the revealed 18 covert conspirators (Krebs, 2002).

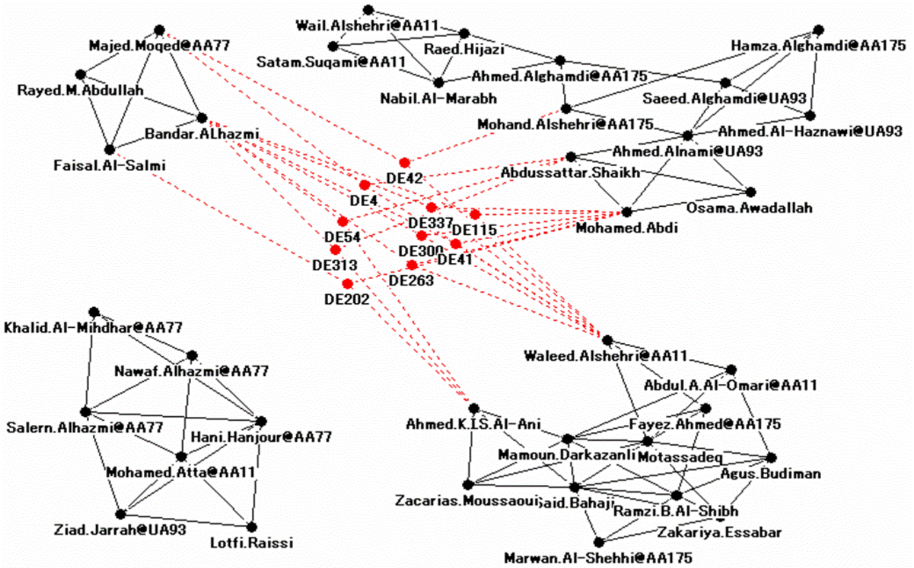

Figure B (corresponding to Fig.10 in attachment I) 4 clusters and 10 highly ranked red nodes corresponding to Mustafa A. Al-Hisawi hidden in the suspicious records. Waleed Alshehri and Mohand Alshehri are retrieved as neighbor persons of red nodes.

d. *Stage 4) Application to business problems*: The method has been applied to two real-business domains (1) redesigning surface inspection system (SIS), a machine for detecting defects on couple charged devices (CCD). (2) Extracting the characteristics of each group and understanding the flow of their particular argument: This has been done by humans taking long time so far, so we aimed to employ the assistance of a computer to understand an argument outline and obtain some materials for educating reasoning skills in mediation or negotiation. Therefore, we applied KeyGraph and Data Crystallization technologies in order to attempt an analysis of long texts in a collaboration between humans and computer based systems. Several researchers have already applied these technologies to analyze speech texts. However, our text logs have following two features; (1) we have several argumentation logs regarding the same issues, (2) we can estimate important issues before the moot arbitration or the moot negotiation by analyzing the case. Here, we extended KeyGraph and the Data Crystallization method to use the above features of our argumentation logs, and achieved our aim.

Significance to the field

The basis of this study has been *chance discovery*, which means to discover a *chance*, defined as an event significant for making a decision. Using existing data in business and natural/social sciences, we have been achieving successful chance discoveries in various domains, including (not restricted to):

- Marketing, where consumer-behaviors from hidden motivations are dealt with,
- Prediction of earthquakes caused by hidden active faults
- Hepatitis treatment, where some observation might be missing in the blood test.

In studies on chance discovery, we have been working well in finding rare but significant events. Data crystallizing means to extend chance discovery to the discovery of significant events which have never occurred in the given data, i.e., from low-frequency to zero-frequency. This means to deal with more uncertain environment where human may miss important event, than we have been dealing with in data mining or chance discovery. We can compare the presented study with previous studies in existing domains, as follows:

1) Empirical studies on terrorist/criminal social networks: Batallas [Batallas, 2006] applied centrality [Freeman, 1979] and brokerage [Cusumano, 2000] to analyzing an aircraft engine development project, and studied the influence of an information leader team, which could be either a bottleneck of information flow or an innovation diffuser. [Keila, 2006] applied factor analysis to studying email exchanges in Enron which lead to the bankruptcy due to the institutionalized accounting fraud. [Klerks, 2002] pointed out that criminal organizations tend to be strings of inter-linked small groups that lack a central leader, but to coordinate their activities along logistic trails and through bonds of friends, and that hypothesis can be built by paying attention to remarkable white spots and hard-to-fill positions in a network. Krebs [Krebs, 2002] investigated the 9/11 terrorist network, and revealed that the relevance of conspirators who reduce the distance between hijackers and enhance communication efficiently. Then, Morselli [Morselli, 2007] investigated Kreb's network from the viewpoint of efficiency and security trade-off, and suggested that more security-oriented structure arises from longer time-to-task of the terrorists' objectives, and that conspirators improve communication efficiency, preserving hijackers' small visibility and exposure. These method search objectively important members and the links between them in the community. The human-centric process of data crystallization, on the other hand, visualizes candidate positions of unobservable leaders in the network, and aims at accelerating the subjective interpretation of analyst, based on his/her/their knowledge.

2) Theories and computational models of complex networks: Complex network, graph theory, and machine learning algorithms help us in obtaining insight on the dynamics of a social network, in addition to summarizing and visualizing a network [Shen, 2007], and analyzing a cognitive network [Krackhardt, 1987]. Scale-free networks [Barabasi, 1999] and small worlds [Watts, 1998] present us much insight on the structure and evolution of a social network: scientists' collaboration, actors in movies etc. A power law in the nodal degree distribution governs the scale-free network. Fenner [Fenner, 2007] proposed an exponential cutoff mechanism to modify the power law. Error attack tolerance [Albert, 2000] and search efficiency [Adamic, 2001] are of particular interest for practical applications. These have been the basis of studies in 1) above, and also of the studies on the visualization part of data crystallization.

3) Evidence extraction and link discovery: A relevant research area to Chance Discovery, where important links of people with other people and with their own actions are to be discovered from heterogeneous sources of data, is Evidence Extraction and Link Discovery (EELD). The difference between Chance Discovery and EELD, at the time we began this project, was in the position of human factors in the research approaches. In Chance Discovery, the visualization techniques such as KeyGraph have been used for clarifying the effect of chances, by enforcing the user's thoughts on scenarios in the real environment. On the other hand, the EELD program mainly contributed to identifying the most significant links among items more automatically and precisely than human. After the one year of this successful project, we showed an improvement of the visualization tool reinforces the process of chance discovery, and this may be regarded as a new feature of the state of chance discovery.

Link discovery has been applied to predicting collaboration between scientists from the published co-authorship [Liben-Nowell, 2004]. Adamic [Adamic, 2003] proposed a technique to infer friends and neighbors from the information available on the web. [Singh, 2004] applied a hidden Markov model and a Bayesian network to predict the behavior of terrorists. Learning of a Bayesian network is extended to study the probabilistic nature of latent variables. Silva [Silva, 2006] studied learning of a structure of a linear latent variable graph. Friedman [Friedman, 1998] studied learning of a structure of a dynamic probabilistic network. The principled analytic approach often suffers from complexity problem. The complexity includes bi-directional and cyclic influence among the many observed and latent nodes (beyond a triad: 1 latent node influencing 2 observed nodes).

Relation to the goal

The sphere of real world applications linked from this basic research is expected to include intelligence analysis aiming to arrest unknown leaders, development of the ideas about new (unknown) products, understanding the hidden but important issues in arbitration or negotiation, etc. We successfully accomplished to show the ability of our methods to solve these new problems, by applying to simulated complex problems and scaled-down simplified versions of these real up-to-date problems.

(5) Personnel Supported: List the professional personnel supported by the contract and/or the personnel who participated significantly in the research effort.

Yoshiharu Maeno, Mr: Developed and implemented the new method human-interactive annealing.

(6) Publications: List peer-reviewed publications submitted and/or accepted during the contract period.
Publications

[1] Miura, T., Katagami, D., and Nitta, K., Analysis of Negotiation Logs using Diagrams, *JSiSE Research Report*, Vol.22, pp.33-38, (2007)

[2] Toshiko Wakaki, Hajime Sawamura, Katsumi Nitta: An Integrated System of Semantic Web Reasoning and Argument-Based Reasoning, Advances in Intelligent Web Mastering (Proceedings of the 5th Atlantic Web Intelligence Conference-AWIC'2007), Advances in Soft Computing 43, Springer, pp. 349-356 (2007).

[3] Takahiro Tanaka, Norio Maeda, Daisuke Katagami, Katsumi Nitta: Characterized Argument Agent for Training Partner, Proceedings of the 1st International Workshop on Juris Informatics (in Conjunction with the 21st Annual Conference of JSAI), pp.30-41 (2007).

[4] H. Sawamura, T. Wakaki, K. Nitta: The Logic of Multi-Valued Argumentation and its Application to Web Technology, The 1st International Conferebce on Computational Models of Argument (COMMA 06), pp.291-296, IOS Press, Liverpool (UK), (2006).

[5] K. Yamamoto, D. Katagami, K. Nitta, A. Aiba, H. Kuwata: The Credibility of Posted Information in a Recommendation System on a Map, The 15th International World Wide Web Conference (WWW2006), pp.985-986 (2006).

[6] T. Tanaka, N. Maeda, D. Katagami, K. Nitta: Characterized Argument Agent for Training Partner, New Frontiers in Artificial Intelligence: JSAI 2007 Conference and Workshops Revised Selected Papers, Lecture Notes on Artificial Intelligence, Springer (2007) (To be published).

[7] H. Akiyama, D. Katagami, K. Nitta: ``Training of the Agent Positioning using Human's Instruction'', Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.11, No.8, pp.998-1006 (2007).

[8] Yoshiharu Maeno and Yukio Ohsawa, 'Analyzing covert social network foundation behind terrorism disaster', *International Journal of Services Sciences* (available e-print http://arxiv.org/abs/0710.4231). (2007)

[9] Yoshiharu Maeno and Yukio Ohsawa, 'Detecting invisible relevant persons in a homogeneous social network', *Lecture Notes in Computer Science*, Vol. 4490, pp.74-81, Springer-Verlag. (2007)

[10] Kenichi Horie, Yoshiharu Maeno and Yukio Ohsawa, Data crystallization applied for designing new products, *Journal of Systems Science and Systems Engineering* Vol. 16, pp.34-49 (2007).

[11] Kenichi Horie, Yoshiharu Maeno and Yukio Ohsawa, Human-Interactive Annealing Process with Pictogram for Extracting New Scenarios for Patent Technology, *Data Science Journal* Vol.6 S132-S136 (2007a)

[12] Maeno, Y., and Ohsawa, Y., 'Catalyst personality for fostering communication among groups with opposing preference', *Lecture Notes in Artificial Intelligence*, Vol. 4570, pp.806-812, Springer-Verlag. (2007)

[13] Yoshiharu Maeno and Yukio Ohsawa, 'Trigger to switch individual's interest toward unconscious preference', *Lecture Notes in Artificial Intelligence,* Vol. 4693, pp.970-977, Springer-Verlag. (2007)

[14] Yoshiharu Maeno and Yukio Ohsawa, Human-Computer Interactive Annealing for Discovering Invisible Dark Events, *IEEE Transaction on Humatronics* Vol.54, No.2, pp.1184 - 1192 (2007)

[15] Yoshiharu Maeno and Yukio Ohsawa, Understanding of dark events for harnessing risk, *Chance Discovery for Real World Decision Making*, Chapter 22, Springer Verlag (2006)

[16] Kenichi Horie, Yukio Ohsawa, Product Designed on Scenario Maps Using Pictorial KeyGraph, *WSEAS Transaction on Information Science and Application*, Vol.3 No.7, pp.1324-1331 (2006)

(7) Interactions: Please list:

(a) Participation/presentations at meetings, conferences, seminars, etc.

Toshiko Wakaki, Hajime Sawamura, Katsumi Nitta: An Integrated System of Semantic Web Reasoning and Argument-Based Reasoning, the 5th Atlantic Web Intelligence Conference-AWIC' (2007).

Takahiro Tanaka, Norio Maeda, Daisuke Katagami, and Katsumi Nitta: Characterized Argument Agent for Training Partner, the 1st International Workshop on Juris Informatics (in Conjunction with the 21st Annual Conference of JSAI), Miyazaki, Japan (2007).

H. Sawamura, T. Wakaki, and Katsumi Nitta: The Logic of Multi-Valued Argumentation and its Application to Web Technology, The 1st International Conference on Computational Models of Argument (COMMA 06), Liverpool (UK), (2006).

K. Yamamoto, D. Katagami, Katsumi Nitta, et al: The Credibility of Posted Information in a Recommendation System on a Map, The 15th International World Wide Web Conference (WWW2006), (2006).

Yoshiharu Maeno, Yukio Ohsawa, and Takaichi Ito: Trigger to switch individual's interest toward unconscious preference, International Conference on Knowledge-Based and Intelligent Information \& Engineering Systems (KES), Vietri sul Mare, Italy, September 2007.

Yoshiharu Maeno Yukio Ohsawa, and Takaichi Ito: Restoring missing network nodes featuring in relevance-to-visibility ratio, Joint Conference on Information Sciences (JCIS), Salt Lake City, USA July 2007.

Yoshiaru Maeno, Yukio Ohsawa, and Takaichi Ito, Catalyst personality for fostering communication among groups with opposing preference, The 20th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE 2007), Kyoto, Japan (2007)

Kenichi Horie, Yoshiharu Maeno, Yukio Ohsawa: Designing New Product Scenarios for Patent by Human-Interactive Annealing with Pictogram, IEEE ICDM Workshop on Risk Mining, Hong Kong (2006)

Yoshiharu Maeno, Kiichi Ito, Kenichi Horie, Yukio Ohsawa: Human-interactive annealing for turning threat to opportunity in technology development, *IEEE ICDM Workshop on Risk Mining,* Hong Kong (2006)

Yoshiharu Maeno, and Yukio Ohsawa Stable deterministic crystallization algorithm for discovering hidden hubs Goda Shinich, and Yukio Ohsawa Chance Discovery in Credit Risk Management, *IEEE Annual Conference on Systems, Man, and Cybernetics*, Taipei, Taiwan (2006)

Yoshiharu Maeno, and Yukio Ohsawa, Hidden structure visualization adaptive to human's prior understanding, *the 9th Joint Conference on Information Sciences (JCIS)*, Kaohshung, Taiwan (2006)

(b) Describe cases where knowledge resulting from your effort is used, or will be used, in a technology application. Not all research projects will have such cases, but please list any that have occurred.

- Visualizing the data of patent lists of a company, with our method of data crystallization, enabled to see new technologies not yet existing in the world. Publications [11] and [12] by Horie et al includes the information about this. See Appendix II in the attachment.

(8) New:

(a) List discoveries, inventions, or patent disclosures. (If none, report None.).

- The basic method of data crystallization, enabling to realize hidden leaders and hidden demands in the market.
- The extension of data crystallization, reflecting human intelligence to the discovery of hidden events, hidden members of a community, hidden concepts underlying existing patents.
- The results of the proposed methods applied to the test data provided by AOARD, and to real data on courtesy and business negotiations.

Patent disclosures: None

(b) Complete the attached **"DD Form 882, Report of Inventions and Subcontractors."**

(9) Honors/Awards: List honors and awards received during the contract period, or emanating from the AOARD-supported research project.

Student Travel Award, for "Designing New Product Scenarios for Patent by Human-Interactive Annealing with Pictogram." IEEE ICDM Workshop on Risk Mining, Hong Kong (2006)

(10) Archival Documentation: This section should include a description of your work at a level of technical detail that you think to be appropriate. Submission of reprints/preprints often satisfies this requirement. If you have questions on how to prepare this section, please discuss this matter with your AOARD program manager. Attached : The following appendixes (I, II, and III)

Appendix I: An Analysis of Terrorist Organization

Appendix II: Corporate R&D Projects aided by Data Crystallization

Appendix III: An Analysis of Mediation and Negotiation Logs by KeyGraph and Data Crystallization